Welcome to Psy 652 Lab!
Module 6:
Merging Datasets, Correlation, Partial Correlation, & Simple Linear Regression (SLR)

Objectives

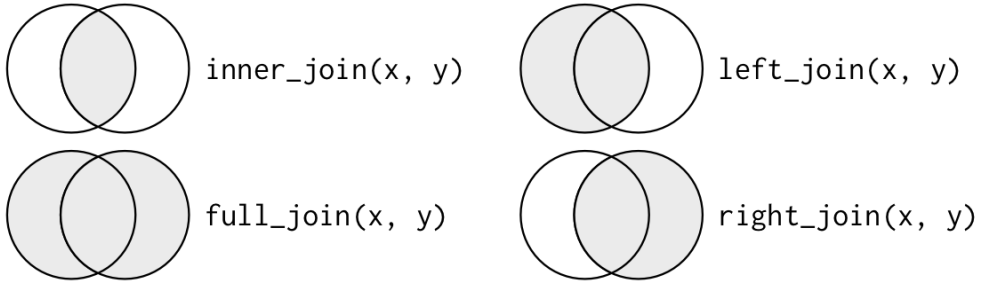
- A quick note on merging datasets (Via *_join())
- Correlations, Partial & Semi-partial correlations
- R tutorial

Merging (Joining) datasets

Joining datasets via tidyverse's _join()

- General format:
  - *_join(left, right, by = "id")
    - * = fill in left, right, inner, or full
    - *left* = left dataset
    - *right* = right dataset
    - *By = "id"* = The id to match on (change to variable you want to match on)

  - left_join() keeps all id's on the left dataset
  - right_join() keeps all id's on the right dataset
  - inner_join keeps only id's present on both dataset
  - full_join keeps every id in both datasets



ID	X1
1	a1
2	a2

ID	X2
2	b1
3	b2

inner_join

ID	X1	X2
2	a2	b1

left_join

ID	X1	X2
1	a1	NA
2	a2	b1

right_join

ID	X1	X2
2	a2	b1
3	NA	b2

full_join

ID	X1	X2
1	a1	NA
2	a2	b1
3	NA	b2

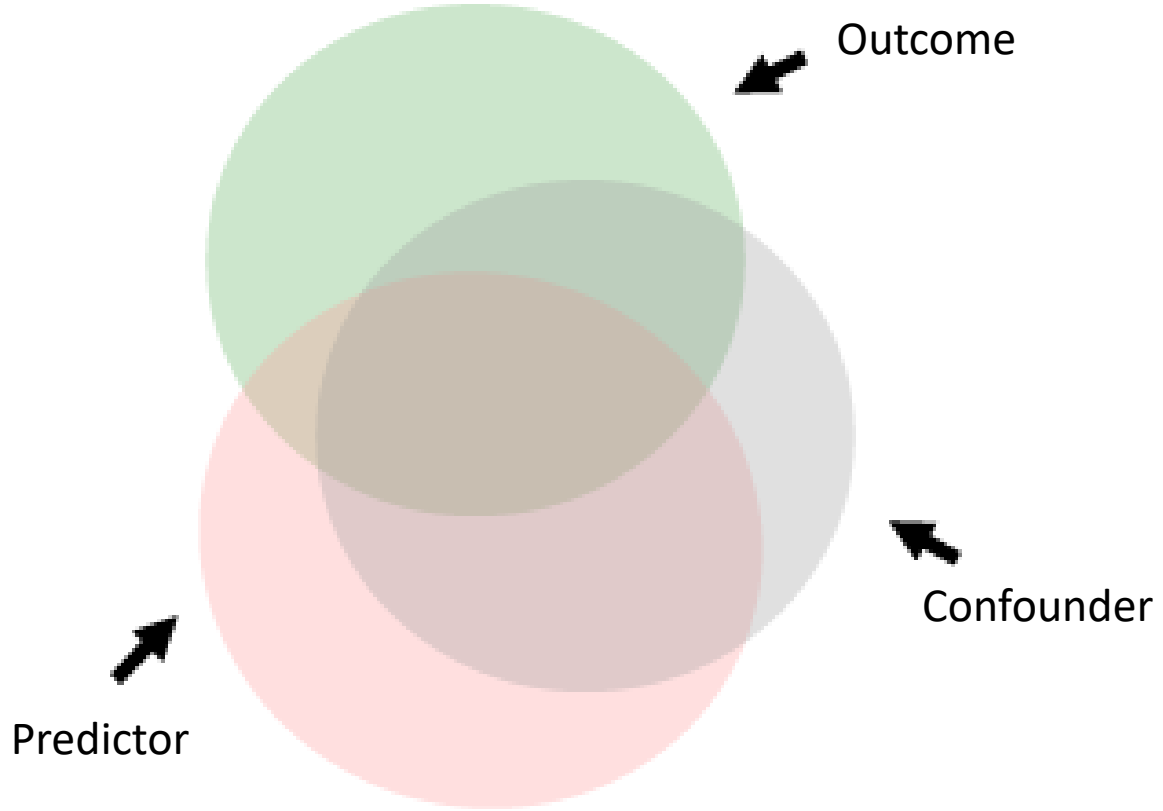
Partial and semi-partial correlations

Partial and semi-partial correlations

- Partial and semipartial correlations are easily confused!
  - Both are correlations that control for a 3rd variable (a confounder)
    - However, the way they control for the variables is different

- **Partial correlation** = Removes part the effect of a confounder from both the predictor AND the outcome
- **Semi-partial correlation** = removes the effect of a confounder from ONLY the predictor variable

Partial and semi-partial correlations

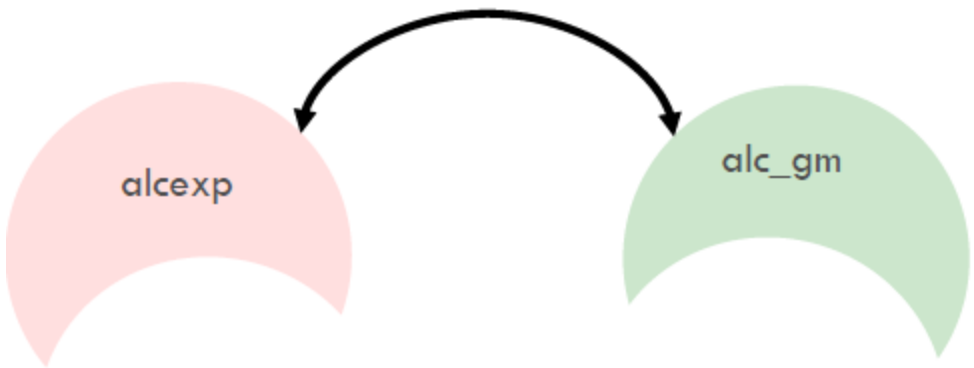


The venn diagram is for illustrative purposes and not drawn to perfect scale.

Thanks to Kim Henry for these visuals

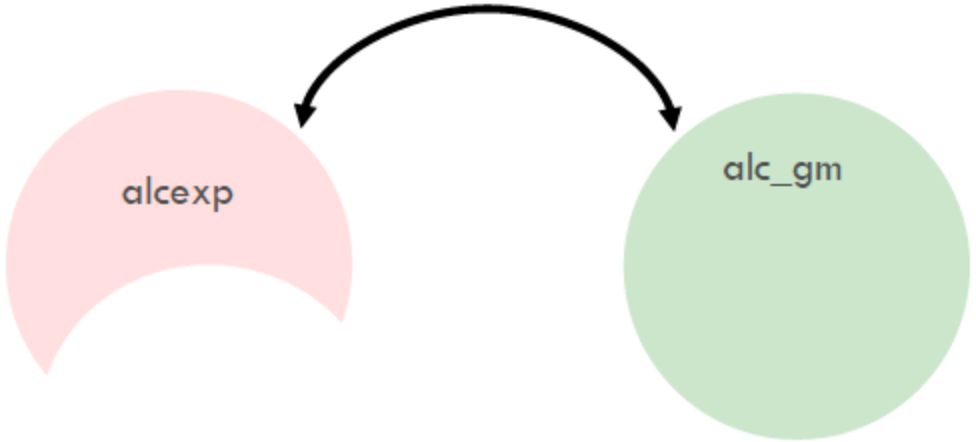
Partial Correlation

For a predictor (alcexp) and outcome (alc_gm) of interest, partial correlation first removes from both alcexp and alc_gm all variance which may be accounted for by the other predictors (in this case, just one, typ_drks), then correlates the remaining variance of alcexp (the residual) with the remaining variance of alc_gm (the residual). Here, the partial correlation between alcexp and alc_gm is .292. Notice that this is the correlation of the two residuals that we obtained in our previous activity.



Semi-Partial (Part) Correlation

For a predictor (alcexp) and outcome (alc_gm) of interest, semi-partial correlation first removes from the predictor (alcexp) all variance which may be accounted for by the other predictors (in this case typ_drks), then correlates the remaining variance of alcexp (the residual) with y. Here the semi-partial (part) correlation between alcexp and alc_gm is .215.



Thanks to Kim Henry for these visuals

Let’s code!